Basics of Markov Chain Monte Carlo (MCMC) and an Introduction to WinBUGS

Michael Anderson, PhD
Hélène Carabin, DVM, PhD

Department of Biostatistics and Epidemiology
The University of Oklahoma Health Sciences Center

May 19, 2016

Outline

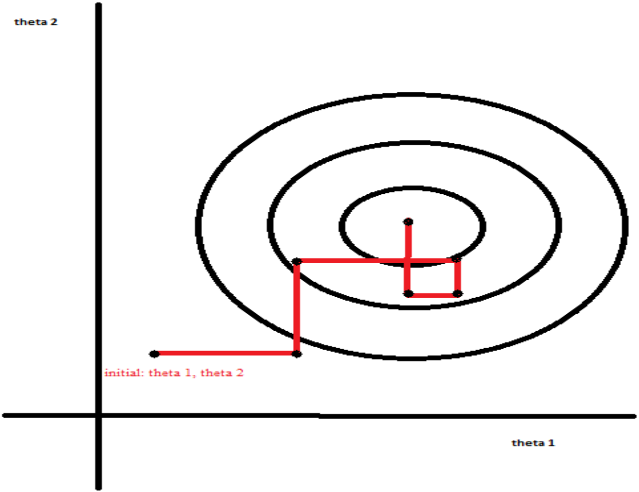
Gibbs Sampling

Suppose we want to obtain samples from $p(\theta_1, \theta_2 | x_1, \ldots, x_n)$. Suppose further that we know $p(\theta_1 | \theta_2, x_1, \ldots, x_n)$ and $p(\theta_2 | \theta_1, x_1, \ldots, x_n)$.
How it works:

1. Choose an initial value for θ_2 say θ_2^0.
2. Obtain θ_1^1 from $p(\theta_1 | \theta_2^0, x_1, \ldots, x_n)$.
3. Obtain θ_2^1 from $p(\theta_2 | \theta_1^1, x_1, \ldots, x_n)$.
4. Repeat steps 2 and 3 with the new θs a large number of times.

Gibbs Sampling

This produces a Markov Chaing that "explores" the parameter space.



Metropolis Algorithm

For the Gibbs sampler we need $p(\theta_1|\theta_2, x_1, \ldots, x_n)$...but often we only have $g(\theta_1|\theta_2, x_1, \ldots, x_n) \propto p(\theta_1|\theta_2, x_1, \ldots, x_n)$
How it works:

1. Pick an arbitrary point for the random walk.
2. Generate a candidate from a symmetric proposal distribution.
3. Compute $r = \frac{g(candidate)}{g(current)}$.
4. $$\text{Let new value} = \begin{cases} \text{candidate with probability min(r,1)} \\ \text{current, otherwise} \end{cases}$$
5. Repeat steps 2-4 a large number of times.

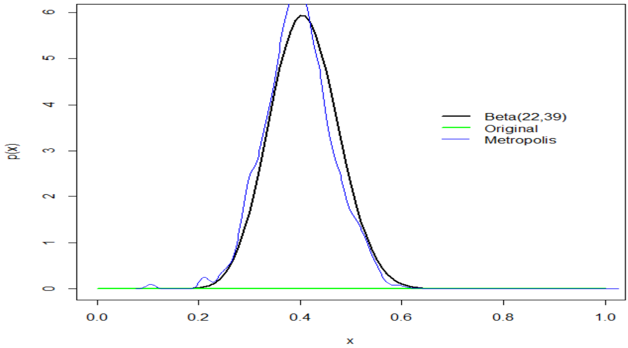
Point: Likelihood and Prior are all we need!

Metropolis Algorithm Example

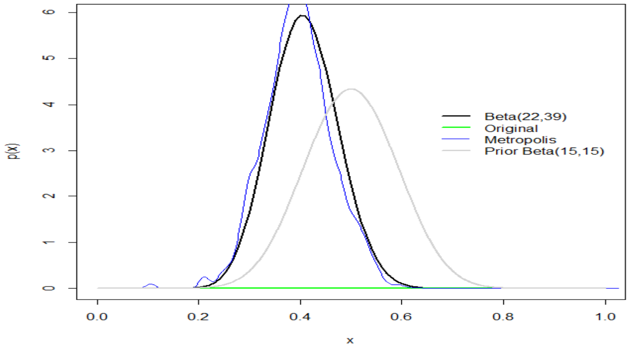
Earlier, we derived the posterior distribution of the success probability from a sample of 24 lung cancer subjects, 7 of which were female. In that example we used our previous knowledge of pdfs to make the integral in the demoninator go to 1. Suppose we want to simply specify the prior and likelihood and employ the Metropolis Algorithm to take care of the rest. Recall

- Prior: $p(\theta) = Beta(15, 15)$
- Likelihood: $p(x1, \ldots, x_n|\theta) = \theta^7(1-\theta)^{24-7}$
- Posterior:$p(\theta|x1, \ldots, x_n) = Beta(15+7, 24-7+15)$

Metropolis Algorithm Example



Metropolis Algorithm Example



Markov Chain Monte Carlo

- In Bayesian analyses, all inference is on $p(\theta|x_1, \ldots, x_n)$
- The vector θ might have many parameters $\theta = (\theta_1, \ldots, \theta_k)$
- Suppose we want $E(\theta_i) = \int \theta_i p(\theta_i|x_1, \ldots, x_n) d\theta_{(-i)}$
- Note: $\theta_{(-i)}$ is the vector θ excluding θ_i.

Now suppose we can draw a random sample from $p(\theta|x_1, \ldots, x_n)$

sample 1 $(\theta_1^{(1)}, \ldots, \theta_k^{(1)})$
sample 2 $(\theta_1^{(2)}, \ldots, \theta_k^{(2)})$
$\vdots$ $\vdots$
sample B $(\theta_1^{(B)}, \ldots, \theta_k^{(B)})$

Note: $\theta_1^{(1)}, \ldots, \theta_1^{(B)}$ is a sample from $p(\theta_1|x_1, \ldots, x_n)$

Monte Carlo Markov Chain

Monte Carlo estimation says that

- $E(\theta_1) \approx \frac{1}{B}\sum_{j=1}^{B}\theta_1^{(j)}$
- $E(\theta_2) \approx \frac{1}{B}\sum_{j=1}^{B}\theta_2^{(j)}$
- $E(g(\theta_1)) \approx \frac{1}{B}\sum_{j=1}^{B}g(\theta_1^{(j)})$

WinBUGS Example

See the files “Intro to WinBUGS.doc” and “Code Intro to WinBUGS.odc”

Compare to Frequentist Approach

Let's briefly compare and contrast the Bayes and Frequentist approaches for this example.

Table: 95% Confidence and Credible intervals for the WinBUGS example.

Method	Prior	Estimate	95% Interval
Frequentist	NA	0.2917	(0.1098,0.4735)
Bayesian	Beta(15,15)	0.4076	(0.2799,0.5416)
Bayesian	Beta(1,1)	0.3071	(0.1489,0.4941)

Interpretation of intervals...
Frequentist: "In repeated sampling from this population, 95% of all intervals constructed in this mannner will contain the true parameter value."
Bayes: "There is a 95% chance the interval contains the true parameter value."